Data-Driven Analysis of Structure Borne Sound for the Detection of Hip-Stem Implant Loosening

Sascha Spors[1], Karl Völzer[2], Rainer Bader[2], and Daniel Kluess[2]
[1] *Institute of Communications Engineering, Universität Rostock, Email: Sascha.Spors@uni-rostock.de*
[2] *Biomechanics and Implant Technology Research Laboratory, Universitätsmedizin Rostock, Germany*

Introduction

A major reason for the revision of total hip replacements is the loosening of the hip-stem [1]. The loosening is often detected too late due to the low sensitivity and specificity of current diagnostic methods [2], e.g. like radiographs, scintigrams and arthrograms. The osseointegration of the hip-stem, as well as the health of the periprosthetic bone are clinically relevant for the structural health of the implant system.

It has alternatively been suggested to analyse the structure borne sound in an hip-stem generated by a forced vibration excitation [3, 4]. The method is based on the assumption that vibration modes of the hip-stem change with the stiffness of the mechanical connection to the femur. In order to enhance the method, an existing total hip-stem was redesigned to house a unit for the excitation of structure borne sound [5, 6]. The system is depicted in Figure 1. It consists of a mechanical oscillator built from a small magnetic ball attached to a spring. The entire unit is mounted inside a hollow cylinder. The mechanical oscillator is excited by an extracorporeal air coil driven electrically at the resonance frequency of the mechanical oscillator. After a short transient phase the ball hits the implant periodically from the inside generating a series of impacts and responses.

In a first series of experiments, the modified hip-stem was mounted in artificial bones. Two different loosenings (small and large bone defect) were considered besides a pressfit. A test rig was built to simulate the static forces acting on the implant. The structure borne sound was picked up by three acceleration sensors mounted at three different positions on the bone. The captured data was analysed and a classificator for the loosening states was developed [7, 8]. It consisted of a principle component analysis (PCA) for dimensionality reduction and a support vector machine (SVM) for classification acting on the magnitude spectrum of the segmented responses.

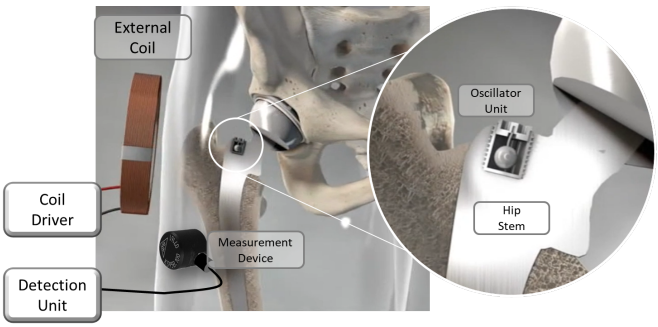


Figure 1: System for the in-situ monitoring of a total hip-stem replacement using structure borne sound.

However, the method requires experimental data from all three states for training of the classifier. In a clinical application, only data from the pressfit state is available directly after implantation. The experimental setup was furthermore very well controlled and the resulting data showed a low amount of inter-individual variability due to the use of artificial bones. The aim of this contribution is to enhance the technique in terms of robustness and applicability in practical situations. For this purpose, the experimental setup is modified and an enhanced data-driven classification which only uses data from the pressfit state is developed and evaluated to detect the loosening.

Experimental Setup

Four human femur bones from body donations (2 female/2 male) are used in the experiments. The bones show a considerable variation in size and shape. The implant with the custom excitation unit was mounted in the bones. Three different states were considered for each of the four bones:

1. pressfit,
2. small loosening, and
3. large loosening.

The latter two by simulating a bone defect removing part of the bone around the implant. The testbed was further extended towards a simple tissue and skin model by placing the bones in a latex cylinder filled with water. The static forces acting on the implant were simulated by a weight. Three acceleration sensors (KS95B from Metra Mess- und Frequenztechnik in Radebeul e.K.) were mounted on the latex cylinder. The signal generation for the air coil and the signal capture for the sensors were performed by a data acquisition interface (USB-6216 from National Instruments) at a sampling rate of $f_\mathrm{s} = 80$ kHz.

The bone with implant was placed in the test rig for the experiments and the air coil was continuously driven by a rectangular signal with a period of $T = \frac{1}{50}$ s. The signals from the acceleration sensors where captured synchronously. A total of 12 experimental series were conducted (4 bones $\times$ 3 conditions).

Data Preparation

Each of the captured signals from the three acceleration sensors contains the responses to a series of impacts of

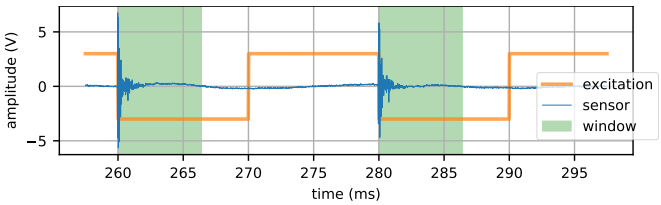


Figure 2: Example: excitation signal of the air coil, signal from the acceleration sensor and window used for extracting the responses.

sample	pressfit	loosening	
		small	large
female 1	33750	32850	29250
female 2	37775	33750	36000
male 1	47334	47975	60340
male 2	51750	33750	33750

Table 1: Number of extracted feature vectors for the different experimental conditions.

the oscillation unit. Exemplary, Figure 2 shows two of these responses (blue line). Inspection of the captured signals revealed that the fundamental frequency of the excitation signal is present. In order to suppress it, the signals were filtered with a band-stop filter having its notch at the fundamental frequency of the excitation signal. The individual responses were then extracted from the filtered signal by a rectangular window of $T \approx 6.4$ ms. Hereby the window is aligned with the rising edge of the excitation signal (green areas in Fig. 2). The windowed responses were then transformed into the discrete frequency domain using a discrete Fourier transform (DFT) for real-valued signals followed by taking the magnitude. In a last step, the magnitude spectra were normalized. This resulted in feature vectors of length $D = 256$ bins. The signal pre-processing is depicted in Figure 3. Using the magnitude spectrum of the individual responses as features is motivated by the rationale that different loosening states result in different vibration modes of the implant and that no exact time-alignment of the responses is required. Table 1 lists the number of extracted feature vectors for the different experimental conditions. A detailed quality check was performed on the extracted feature vectors. Figure 4 shows a statistical analysis of the three different states for one of the bones (female 1). The median, and 5% and 95% percentiles of the magnitude spectrum have been computed and plotted. The variation between the features for one condition is clearly visible, as well as the variations between the conditions. We aim at an approach which is invariant to the variations in the pressfit state but can reliably detect systematic deviations from this state.

Anomaly Detection

The aim of anomaly detection techniques is to detect data that significantly deviates from the majority of given normal data instances. Here, the detection of hip-stem implant loosening using data from the pressfit state only constitutes a semi-supervised anomaly detection problem. The pressfit data represents the normal state, whereby the data from the loose states should be detected as anomaly. Anomaly detection is a challenging problem because only (part of) the normal data is known and only few or none of the anomalies. Deep learning has advanced the task significantly. In [9] three different classes of approaches have been identified: anomaly detection using (i) generic feature extraction, (ii) a learned representation of normality and (iii) end-to-end anomaly score learning. We use an approach from the second class which is detailed in the following.

Model Architecture

We aim to find a mapping $\mathcal{M}$ of the features $\boldsymbol{x}_n$ onto themselves which captures the key regularities of the pressfit (normal) state. The underlying rationale is, that the features from the loose states (anomalies) cannot be reconstructed well by this mapping. Consequently, a loose state shall be detected by inspecting the reconstruction error between the feature $\boldsymbol{x}_n$ and its reconstruction $\hat{\boldsymbol{x}}_n = \mathcal{M}\{\boldsymbol{x}_n\}$ using a scoring function $e(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n)$. If the scoring function exceeds a chosen threshold $\epsilon < e(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n)$ an unseen feature is classified as belonging to the loose state. We have chosen the mean absolute error (MAE) as scoring function

$$e(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{1}{D} \sum_{d=0}^{D-1} |x_d - \hat{x}_d| \,, \tag{1}$$

where x_d denotes the d-th element of the vector $\boldsymbol{x}$. The threshold ϵ is chosen as the 99% percentile of the MAE distribution.

We use a deep convolutional autoencoder for the mapping $\mathcal{M}$. Autoencoders are a generic feature learning techniques which have been used in a wide variety of applications. For instance for coding, dimensionality reduction, denoising, and anomaly detection. Autoencoders aim to learn an efficient encoding of the features into a lower-dimensional latent space. The encoding is validated and refined by attempting to reconstruct (decode) its input from the latent space. This way, the autoencoder learns a representation for the given set of features by training the network to ignore insignificant data, e.g. noise.

Figure 5 illustrates the architecture of the autoencoder. It was chosen by pre-evaluating various sensibly chosen variants. The autoencoder consists of two-layers without skip connections. In the encoding branch, each layer is constructed from a one-dimensional convolutional neural network (CNN) followed by batch normalization. In between the layers a one-dimensional strided CNN is used for downsampling followed by a dropout layer for regularization. The decoding branch is symmetric in structure but uses one-dimensional transposed strided CNNs for upsampling and a final one-dimensional CNN for feature reconstruction. The weights of the CNNs were trained by the Adam optimizer using the mean squared error (MSE) loss function. Early stopping was used to prevent overfitting. The number of convolution kernels, their size, the number of strides and the dropout rate is evaluated as part of the model selection process outlined in the next

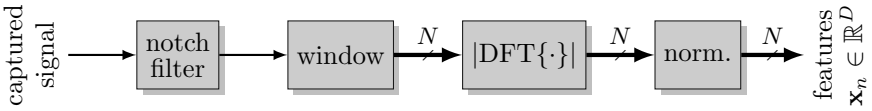


Figure 3: Signal pre-processing used to extract the feature vectors.

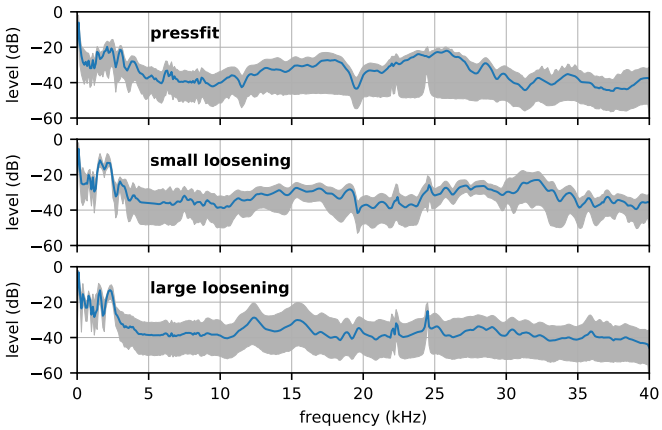


Figure 4: Magnitude spectra of the responses captured for one of the bones (female 1). The median (blue line), and the 5% and 95% percentiles (gray area) are shown.

section.

Model Selection

We aimed a total of five final models, four trained on the data from each of the individual bones separately and one using the combined data from all bones. The latter one should provide insights into the development of a non-individual model. The models were only trained on the features from the pressfit state. The model selection was performed in two stages:

1. 50 model candidates with randomly chosen hyperparameters were trained and
2. the trained models were evaluated with respect to their capability to compactly represent the normal state.

Additionally to the above listed hyperparameters, the learning rate was also included in the model selection. For each of the model candidates, all five hyperparameters were randomly chosen from a set of predefined choices. Then the MAE was computed for each of the candidates and represented as a histogram. The set of hyperparameters that resulted in the most compact distribution of the MAE was chosen for the final model. The final models were then retrained and the thresholds ϵ derived.

Results

In order to evaluate the five final anomaly detection models, the feature vectors $\hat{\boldsymbol{x}}_n$ were reconstructed for the entire respective dataset for all three classes (pressfit, small loosening, large loosening). From this, the MAEs were computed and the feature vectors were classified into belonging to normal or anomaly states using the threshold. The evaluation of the models is based on the MAE histograms and the classification metrics.

model	pressfit		loosening	
	precision	recall	precision	recall
female 1	1.00	0.99	0.99	1.00
female 2	1.00	0.99	0.99	1.00
male 1	1.00	0.99	1.00	1.00
male 2	1.00	0.99	0.99	1.00
combined	0.94	0.99	0.99	0.96

Table 2: Classification performance of the five final models.

Figure 6a shows the MAE histograms using the final model and data for the bone female 1. The pressfit state is represented by a compact MAE distribution. The MAEs of the small and large loosening are significantly higher. Hence, a clear separation between the classes is achieved by the model. This is also reflected by the classification metrics reported in Table 2. The classification performs overall very well. The 1% lower recall rate for the pressfit state is a result of choosing the 99% percentile as detection threshold. For the other three models only the classification metrics are reported (see Table 2).

Figure 6b shows the MAE histograms using the final model and data for the combined case. The distribution of the MAE for the pressfit state is not as compact as for the model female 1. There is also an overlap in MAEs of the pressfit state and the large loosening observable. A reason could be that the two-layer model architecture is not complex enough to reflect the variety in the combined data. The classification performance reported in Table 2 is nevertheless still quite good.

Conclusions

We presented an approach to the data-driven detection of hip-stem implant loosening based on the analysis of structure borne sound and deep anomaly detection. The structure borne sound was generated by a custom made excitation unit magnetically excited by an extracorporeal air coil. The approach aims at extracorporeal in-vivo monitoring of hip-stem implants complementing established diagnosis methods. The results show that loosening can be detected almost certain using individual models and with high accuracy using a non-individual model trained with data from multiple individuals. A cross validation of the non-individual model has not been performed so far. It has to be noted that the experimental setup generated very reproducible measurements and more variations can be expected in a practical application. For instance long-term changes due to ageing or the loss of weight. Future work should include the modelling of these and other aspects, as well as alternative outlier

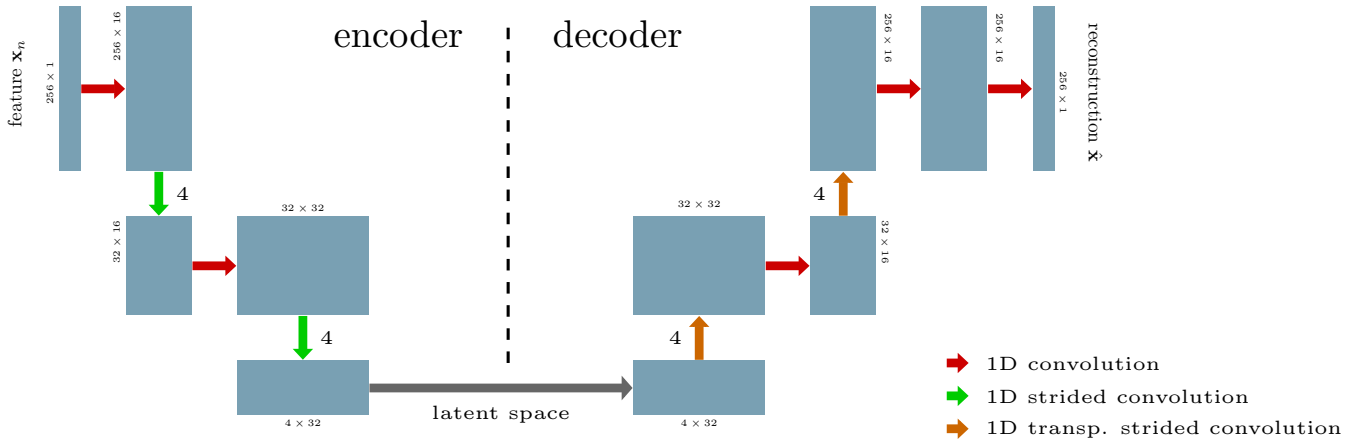
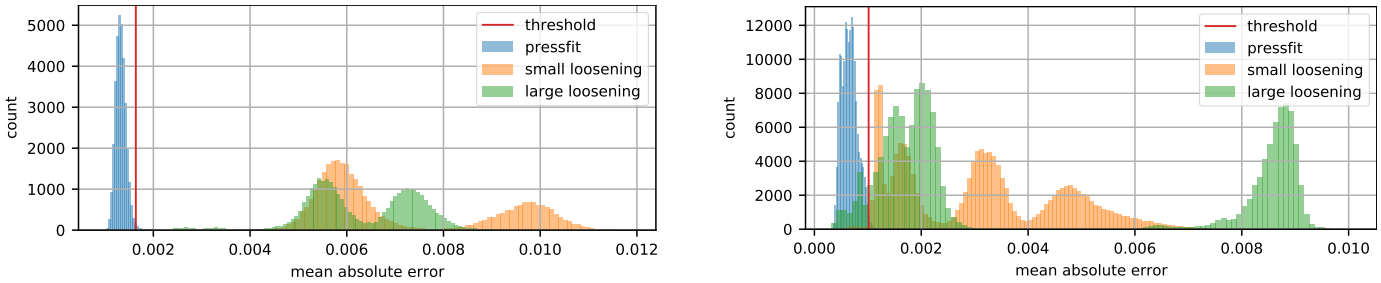


Figure 5: Architecture of the autoencoder used to represent the mapping of the normal features. The shape of the boxes indicates the size of the (intermediate) data representations, the arrows the operations applied to the data.



(a) model and data of female 1

(b) combined model and data from all bones

Figure 6: Histograms for the MAE for three different loosening states for one individual and the non-individual model. The threshold ϵ for detection of a loose state is shown by the red line.

detection strategies and thresholding techniques. In the mid-term we plan to integrate the sensing and detection into an electrically active implant.

Open Science – The implementation of the model, the trained models, and the classifier with some examples are available as data publication [10].

Acknowledgements – The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — SFB 1270/1 – 299150580. And by the European Union Research Council, 7th Framework Program, SMART-HIP, Grant agreement ID 606335.

References

[1] H. Malchau, P. Herberts, T. Eisler, G. Garellick, and P. Soederman, "The swedish total hip replacement register." *The Journal of bone and joint surgery. American volume*, vol. 84-A Suppl 2, pp. 2–20, 2002.

[2] C. Ruther, C. Schulze, A. Böhme, H. Nierath, H. Ewald, W. Mittelmeier, R. Bader, and D. Kluess, "Investigation of a passive sensor array for diagnosis of loosening of endoprosthetic implants," *Sensors (Basel, Switzerland)*, vol. 13, pp. 1–20, 12 2012.

[3] A. Georgiou and J. Cunningham, "Diagnostic accuracy and interobserver variability of plain radiography, subtraction arthrography, nuclear arthrography, and bone scintigraphy in the assessment of aseptic femoral component loosening," *Clinical Biomechanics.*, no. 16, pp. 315–323, 2001.

[4] S. Jaecques, I. Denayer, G. Van der Perre, L. Borgward, J. Larsen, and A. Borgwardt, "Detection of hip stem loosening by vibration analysis: a per-operative case study," Oct. 2002.

[5] T. Bender, M. Sass, W. Mittelmeier, R. Bader, and D. Kluess, "Finite element analysis into eigenfrequencies of a total hip stem with different levels of loosening," in *COMSOL Conference*, Oct. 2015.

[6] T. Bender, M. Sass, R. Bader, and D. Kluess, "SMART-HIP - Entwicklung und Auslegung eines Sensorsystems für die Messung der Verankerung von Implantaten," in *Dresdner Sensor-Symposium*, Dec. 2015.

[7] T. Bender, M. Sass, R. Bader, D. Kluess, and S. Spors, "Employing a principle component analysis and support vector machines to classify different levels of loosening of a total hip system," in *Annual Meeting of the Orthopaedic Research Society*, March 2016.

[8] ——, "Körperschallanalyse zur Klassifizierung einer Lockerung in Hüftendoprothesen," in *German Annual Conference on Acoustics (DAGA)*, March 2016.

[9] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, Mar. 2021. [Online]. Available: https://doi.org/10.1145/3439950

[10] S. Spors, K. Völzer, R. Bader, and D. Kluess, "Model for the data-driven analysis of structure borne sound for the detection of hip-stem implant loosening," September 2021, https://doi.org//10.5281/zenodo.5510331.